第 38 卷第 3 期
录用定稿

计算机应用研究
Application Research of Computers

ChinaXiv合作期刊
Vol. 38 No. 3
Accepted Paper

基于卷积神经网络的语义分割算法研究 *

熊　炜 [1,2†]，童　磊 [1]，金靖熠 [1]，王传胜 [1]，王　娟 [1]，曾春艳 [1]

(1. 湖北工业大学电气与电子工程学院，武汉 430068; 2. 美国南卡罗来纳大学计算机科学与工程系，南卡 哥伦比亚 29201)

摘　要：针对语义分割中残差网络并不能完好地提取图像信息和分割效果差的问题，提出一种联合特征金字塔模型(JFP)用来融合残差网络的输出特征，并结合暗黑空间金字塔池化模型(ASPP)进一步提取特征，在解码部分，应用简单的解码结构，恢复图像尺寸完成语义分割，同时引入注意力模型作为辅助语义分割网络，辅助神经网络进行训练。该方法分别在 Pascal VOC 2012 数据集和增强的 Pascal VOC 2012 数据集对网络进行训练，并在 Pascal VOC 2012 的验证集上进行测试，其平均交并集之比(mIoU)分别达到了 78.55%和 80.14%，表明所提方法具有良好的语义分割性能。

关键词：图像语义分割；联合特征金字塔模型(JFP)；暗黑空间金字塔模型(ASPP)；注意力模型

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2019.12.0705

Research on semantic segmentation algorithm based on convolutional neural network

Xiong Wei[1,2†], Tong Lei[1], Jin Jingyi[1], Wang Chuansheng[1], Wang Juan[1], Zeng Chunyan[1]

(1. *School of Electrical & Electronic Engineering*, *Hubei University of Technology*, *Wuhan* 430068, *China*; 2. *Dept. of Computer Science & Engineering*, *University of South Carolina*, *Columbia*, *SC* 29201, *USA*)

Abstract: In order to solve the problem that the residual network can not extract image information well and the segmentation effect is poor in semantic segmentation, this paper proposed a joint feature pyramid model (JFP) to integrate the output features of the residual network, and then further extract the features in combination with the atrous spatial pyramid pooling module (ASPP) . In the decoding part, this paper applied a simple decoding structure to recover the image size to complete the semantic segmentation This paper also used attention module as the auxiliary semantic segmentation network to assist the training of the neural network. This method trains the network in the Pascal VOC 2012 data set and the enhanced Pascal VOC 2012 data set respectively, and tests it on the verification set of Pascal VOC 2012. The average ratio of intersection and Union (Miou) is 78.55% and 80.14% respectively, which shows that proposed method has good semantic segmentation performance.

Key words: image semantic segmentation; joint feature pyramid module (JFP); atrous spatial pyramid pooling module (ASPP); attention module

0　引言

图像的语义分割是对图像进行像素级别的分割，需要对图像的每一个像素从语义上进行分类[1,2]，同一类别的像素分成同一类别标签，体现在分割结果上就是同一类别的物体属于同一个颜色标签，而不同颜色就是不同类别的物体。

卷积神经网络(CNN)的应用使得图像语义分割快速得到发展，各种基于卷积神经网络的语义分割网络结构被提出。加州大学伯克利分校的 Long 等人[3]提出的完全卷积网络(FCN)，去掉了 CNN 末端使用的全连接层[4]，使得网络最后生成的不是固定的特征向量，而是可以变换尺寸的特征图像，最后进行逐像素的分类以达到语义分割的目的，类似 FCN[5]的思路贯穿在语义分割的研究当中。

FCN 之后，Badrinarayanan 等人[6]提出了 SegNet，用于图像语义分割，是一种深度卷积编码解码架构[7]，并跟随着一个像素级别的分类层，编码过程通过池化逐渐减少位置信息并提取图像更深层的特征，这个过程逐渐缩减输入图像的空间维度，而译码过程会逐渐恢复位置信息，并恢复原有空间维度和对图片进行分割，改进编解码模型[8]也是一些研究的方向。

由于语义分割是逐像素的分类过程，卷积操作使得网络的参数量变大，常常需要加入池化层[9]对图像进行降维处理，以减少参数，这又会产生图像信息丢失的问题，而进行语义分割必须要保持与原图像的像素对齐，每个像素的信息都有意义，这是语义分割面临的最大问题。继而 Yu 等人[10]提出了膨胀卷积，又称空洞卷积[11]，通过这个卷积操作聚合更大尺度的信息，同样的卷积核尺寸，空洞卷积有更大的感知域，有效地解决了语义分割中信息丢失的问题。

此外，如何设计一个神经网络模型也是语义分割研究中的主要内容。金字塔池化模型(ASPP)[11,12]通过应用几个不同核心尺寸的空洞卷积层来扩大感知域，得出不同尺度的特征图，结合并转换成固定大小的特征图，有效地提取了图像的空间尺度信息，不过却增加了网络模型的大小，而 Li 等人[13]引入注意力机制，重新设计了一种注意力金字塔模型(PAN)，进一步提取语义信息，效果进一步提升。Yu 等人[14]从网络层次的信息聚合出发，详细介绍并总结了不同网络层的连接方

收稿日期：2019-12-11；**修回日期：**2020-05-19　**基金项目：**国家留学基金资助项目(201808420418)；国家自然科学基金资助项目(61571182，61601177)；湖北省自然科学基金资助项目(2019CFB530)

作者简介：熊炜(1976-)，男(通信作者)，湖北宜昌人，副教授，硕导，博士，主要研究方向为数字图像处理和计算机视觉(xw@mail.hbut.edu.cn)；童磊(1993-)，男，湖北鄂州人，硕士研究生，主要研究方向为图像处理和语义分割；金靖熠(1994-)，男，湖北武汉人，硕士研究生，主要研究方向为计算机视觉和室内定位；王传胜(1993-)，男，江苏南通人，硕士研究生，主要研究方向为视频去抖动；王娟(1983-)，女，河北邯郸人，讲师，硕导，博士，主要研究方向为人工智能和图像处理；曾春艳(1986-)，女，湖北武汉人，副教授，硕导，博士，主要研究方向为信号处理与压缩感知.

式，再此之上设计了迭代深度聚集(IDA)和分层深度聚集(HAD)模型，通过 IDA 和 HAD 的组合连接，可以设计不同深度和不同连接的深层聚集(DLA)模型进行语义分割，同时这种方法可以加入其他模型中，完成不同的任务。

目前这些网络模型绝大部分都采用了深度卷积神经网络(DCNN)作为骨架网络，在此基础上设计针对图像语义分割的特定神经网络模型，但是 DCNN 也不能完整的提取图像的特征，存在信息丢失的问题，为解决这个问题，本文以 ResNet101 为骨架网络，设计了联合特征金字塔(JFP)模型，结合文献[11, 12]中的 ASPP 模型作为编码结构，以更加完整的提取图像的特征，建立一个简单的解码结构恢复图像信息，为训练本文的语义分割网络模型，损失函数使用 SoftMax CrossEntropy，激活函数使用 ReLu 函数，设计了一个注意力模型作为辅助网络，使网络能更快速的收敛，最终组成本文的图像语义分割方法，提高语义分割的性能。

1　本文方法介绍

1.1　总体框架

本文的总体框架如图 1 所示，首先选择 ResNet101 作为骨架网络进行特征提取，提出了一个 JFP 模型将 ResNet101 输出的后三层进行联合，完善 ResNet101 对特征的提取，解决图像信息丢失的问题；然后将 JFP 的输出接入 ASPP 模型进一步提取图像的空间尺度信息，这部分作为编码结构能够更好的对图像信息进行提取；最后应用简单的解码结构将神经网络的输出图像恢复为原始大小，完成对图像的语义分割；同时，本文设计了一个注意力模型作为辅助语义分割网络，将这个模型的损失函数与语义分割网络的损失函数结合，辅助网络进行训练，提升训练模型的效果。

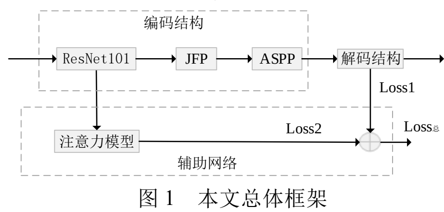


图 1　本文总体框架

Fig. 1Overall framework

1.2　联合特征金字塔模型(JFP)

本文提出的 JFP 模型如图 2 所示。首先，ResNet101 输出的最后三层分别为 block1、block2 和 block3，这三层的输出尺寸减半而深度增加一倍，将这三层的输出分别通过一个卷积，其中，卷积核尺寸为 3，激活函数为 ReLu，在图像边界进行 1 个像素值为 0 的填充，从而不改变输出图像的尺寸，加入批量归一化处理，采用 dropout 为 0.3 以防止过拟合，卷积核的深度为 512，使得输出的深度都变成 512，然后分别通过空洞卷积率为 1、2 和 4 的 3×3 卷积，其中像素填充分别与空洞卷积率相同，不采用 dropout 处理，其中空洞卷积率为 2 和 4 的卷积输出还要在图像边界加入 2 和 4 个像素值为 0 的填充，保持输出尺寸与输出的相同，加入双线性插值，对这两个输出进行调整，使得输出的尺寸与 block1 相同，由 ResNet101 的三层输出经过不同的卷积处理得到三个尺寸与深度相同的输出，与 block1 层的输出尺寸与深度相同，最后将这三个输出与 block1 的输出相加，因此 JFP 模型输出的特征图尺寸与 block1 的输出相同，而深度为 2048。本文在 JFP 模型中使用的空洞卷积率较小，是考虑图片特征能更好的提取，它的感受域提升并不大，模型要比采用大的空洞卷积率的模型要小，但是却十分有效。

1.3　暗黑空间金字塔模型(ASPP)

本文在 JFP 模型后使用 ASPP 模型[11, 12]进一步对图像特征进行处理，其模型结构如图 3 所示。模型输出是由五个相同尺寸和深度的特征图相加得来，将 JFP 模型的输出作为输入，首先，应用 1×1 的卷积，将 JFP 的输出深度降为 256，生成一个尺寸为(h,w)深度为 256 的特征图；其次，应用空洞卷积率为 6，8 和 10 的空洞卷积，在图像边界进行 6、8 和 10 个像素值为 0 的填充，不改变图像尺寸，输出三个尺寸为(h,w)深度为 256 的特征图；然后，应用全局池化结合 1×1 卷积，然后使用双线性插值法恢复图像尺寸，输出一个尺寸为(h,w)深度为 256 的特征图；最后，由这 5 个输出特征图相加得到与 JFP 的输出特征图尺寸相同、深度为 1280 的输出。其中，卷积的激活函数为 ReLu，加入了批量归一化处理。这部分模型应用的空洞卷积率相较于文献[11, 12]减小了，目的是为了减小模型结构，相比于本文 JFP 模型，采用相对大的空洞卷积率，较大的增加了感受域的大小，ASPP 在本文 JFP 的基础上进一步提取图片的空间尺度信息，能更好地提升特征提取的效果。

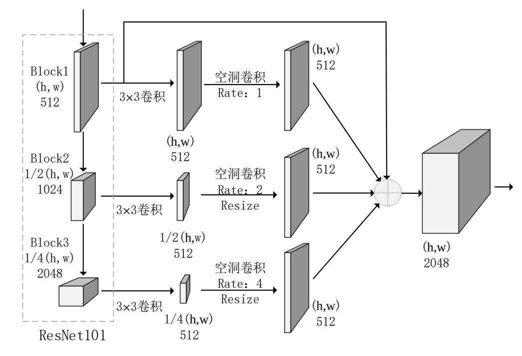


图 2　联合特征金字塔模型(JFP)

Fig. 2　Joint feature pyramid module(JFP)

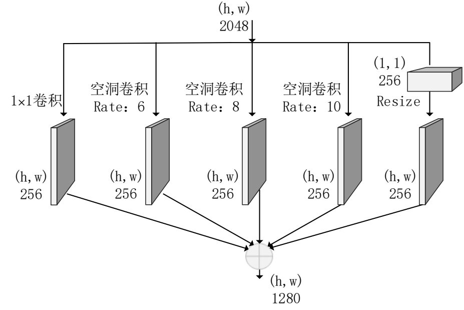


图 3　暗黑空间金字塔模型(ASPP)

Fig. 3　Atrous spatial pyramid module (ASPP)

1.4　解码结构

本文设计了一个简单的解码结构，如图 4 所示，采用 1×1 卷积、3×3 卷积和 1×1 卷积的组合，第一个卷积将输入的深度降为 256，第二个卷积作进一步特征处理，第三个卷积将深度降为 21，与 Pascal VOC 2012 数据集的类别数相同(包括背景)，最后通过双线性插值法将图像尺寸变为 400×400，这个尺寸是数据集裁剪的尺寸，与最开始输入神经网络的图像尺寸保持相同。卷积的激活函数为 ReLu，加入批量归一化处理，而其中 3×3 卷积加入了 0.1 的 dropout，与 JFP 模型中 dropout 的值不同，因为设置不同的 dropout 可以得到更好结果。

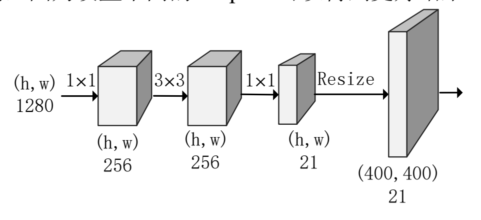


图 4　解码结构

Fig. 4　Decoding structure

1.5　注意力模型

本文设计了一个注意力模型作为语义分割模型的辅助网

络，其结构如图 5 所示，首先将 ResNet101 的 Block2 的输出做一个 1×1 卷积处理，将特征图输出深度降为 21，然后进行全局池化处理，其中卷积过程的激活函数为 ReLu，加入批量归一化处理，最后通过双线性插值法将输出图像尺寸变为 400×400。

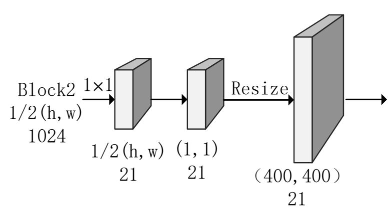


图 5　注意力模型

Fig. 5　Attention module

这一部分的网络是作为辅助网络的，将这个网络的损失函数作为语义分割模型损失的一部分，辅助本文设计的神经网络进行训练，如图 1 中所示，语义分割网络的损失为 *Loss*1，辅助网络的损失为 *Loss*2，*Loss*1 和 *Loss*2 均为 SoftMax CrossEntropy 损失函数所定义，为式(1)所示。

$$Loss = -\sum_{i=1}^{j} y_i' log(y_i) \tag{1}$$

其中，$i, j \in \{1,2,3\cdots,21\}$，$y_i'$ 为标签图像中属于第 i 个类的概率值，即真实分布值，y_i 为语义分割模型输出预测属于第 i 个类的概率值，即预测分布值，y 由 SoftMax 函数定义，如式(2)所示。

$$y_i = softmax(x_i) = \frac{exp(x_i)}{\sum_{j} exp(x_i)} \tag{2}$$

最终训练网络的损失函数为 $Loss_{总}$，其关系式如式(3)所示。

$$Loss_{总} = Loss1 + 0.5 \times Loss2 = -\sum_{i=1}^{j} y_i' log(\frac{exp(x_i^1)}{\sum_{i=1}^{j} exp(x_i^1)}) - 0.5 \times \sum_{i=1}^{j} y_i' log(\frac{exp(x_i^2)}{\sum_{i=1}^{j} exp(x_i^2)}) \tag{3}$$

其中，0.5 为本文设置辅助网络对整个模型损失函数的影响系数。在语义分割网络模型的卷积层中选择 ReLU 函数作为激活函数，最后层使用 SoftMax CrossEntropy 损失函数进行分类，这样简单而且高效。

2　实验结果与分析

作者将本文提出的方法与近 3 年的方法进行了大量对比实验。本文使用的数据集来源于 Pascal VOC 2012 数据集，有两种类型，第一种包括 1464 张训练图像，1446 张验证图像和 1456 张测试图像；第二种增强数据集，加入了 Pascal 边界检测数据集进行扩充，包括 10582 张训练图像，1446 张验证图像和 1456 张测试图像。数据集图像分辨率大小 300-500 不等，在验证集上进行测试，标签图片包括背景类总共有 21 个不同类别，使用不同的颜色表示不同类别的物体，其中训练集和验证集的标签图像是公布的，因此本文在训练集上进行网络训练，在验证集上进行指标评价，在测试集上比较语义分割结果。

本文使用 Pytorch 作为深度学习框架，建立语义分割模型，GPU 型号为 8G GeForce RTX 2070，使用平均交并集之比(mIoU)作为性能评估指标，mIoU 值越高表示语义分割效果越好。实验中，ResNet101 骨架网络使用的是在 ImageNet 上进行预训练的参数，将输入图片大小调整为 400×400，然后裁剪为 384×384(预处理)，设置迭代周期为 180，batchsize 为 8，学习率为 0.001，学习率衰减为 0.9，权重衰减为 0.0001。在上述两种 Pascal VOC 2012 数据集上都进行了实验，首先，在 Pascal VOC 2012 数据集(1464 张训练图片)上对网络进行训练，然后在增强的 Pascal VOC 2012 数据集(10582 张训练图片)上对网络进行训练，并都在 Pascal VOC 2012 数据集的验证集上进行测试，测试结果如表 1 所示，本文 mIoU 值分别为 78.55%和 80.14%，其他为文献[11~13, 15~22]中的方法在验证集中的评价结果，这些方法的骨架网络均在 ImageNet 上进行预训练，可见本文的方法在使用 10582 张训练图像的时候，mIoU 超过了其中的方法，而使用 1464 张训练图像的时候，mIoU 超过其中一些方法或者与其中一些方法接近。

表 1　Pascal VOC 2012 验证集 mIoU 结果

Tab. 1　Miou results of Pascal VOC 2012 validation set

方法	骨架网络	mIoU/%
BlitzNet[15]	ResNet101	72.40
LadderDenseNet[17]	DenseNet	78.01
Context+Decoder+CRFs[22]	ResNet101	75.26
MsNet-4[21]	ResNet101	75.80
DFN[20]	ResNet101	79.54
DeeplabV3[11]	ResNet101	78.51
DeeplabV3+[12]	ResNet101	78.85
PAN[13]	ResNet101	79.38
SDN[16]	DenseNet	78.60
Auto-Deeplab[18]	ResNet101	75.26
DUpsamling[19]	Xception	79.67
本文 1(1464 张训练图片)	ResNet101	78.55
本文 2(10582 张训练图片)	ResNet101	**80.14**

本文方法的速度对比如表 2 所示，使用 1464 张训练图像时，本文方法训练时间比 DeeplabV3+多 0.69 小时，验证速度比 DeeplabV3+慢 1.79 个帧率(fps)，测试速度慢 1.13fps，使用 10582 张训练图像时，本文方法训练时间比 DeeplabV3+多 5.01 小时，验证速度比 DeeplabV3+慢 1.56 个帧率(fps)，测试速度慢 1.37fps，本文方法实现语义分割的速度下降近 1.2fps，即本文方法增加了网络模型的计算量，使得语义分割的速度有一定下降，不过却并不，却显著提升了语义分割的效果，在计算量与效果的综和比较上，体现了本文语义分割方法的优越性。

表 2　本文方法速度对比

Tab. 2　Speed comparison of our method

方法	mIoU/%	train/h	val/fps	test/fps
DeeplabV3+(1464)	77.59	**16.16**	**22.36**	**17.13**
Ours1(1464)	**78.55**	16.85	20.57	16.00
DeeplabV3+(10582)	78.89	**117.45**	**22.29**	**17.20**
Ours1(10582)	**80.14**	122.46	20.73	15.83

本文复现了 DeeplabV3+[12]中的方法，并在每个类别的交并集之比(IoU)上与本文的方法进行比较，结果如表 3 所示，可见在大部分类别的物体分割上，本文的方法具有明显的提升，而在使用较少训练图片的时候，mIoU 值接近 DeeplabV3+[12]使用 10582 张训练图片的结果，而在有些物体上，语义分割效果有下降，可能是遮挡、光照和物体细节上的差异导致。

另外分别使用本文方法和 DeeplabV3+[12]方法在 Pascal VOC 2012 测试集上生成语义分割结果，如图 6 所示，其中，a 为原图，b 为 DeeplabV3+[12]使用 1464 张训练图像的语义分割结果，c 为本文方法使用使 1464 张训练图像的语义分割结果，d 为 DeeplabV3+[12]使用 10582 张训练图像的语义分割结果，e 为本文方法使用 10582 张训练图像的语义分割结果。可见，对于物体的大部分区域，与 DeeplabV3+[12]方法比较分割效果显著，而小的细节上都存在一定的忽视和错误。而本文的方法最终在物体整体语义分割效果上具有一定提升，错误的语义分割也较少，分割出了其他类别的物体，或将其他物体分割为目标类别。

表 3 Pascal VOC 2012 验证集 21 类 IoU 结果

Tab. 3 21 classes iou results of Pascal VOC 2012 validation set

类别	1464 training images		0582 training images	
	DeeplabV3+	本文 1	DeeplabV3+	本文 2
background	92.38	93.88	93.90	**94.26**
aeroplane	86.26	89.76	90.29	**93.51**
bike	**62.27**	62.07	47.55	46.15
bird	87.14	88.14	89.17	**90.68**
boat	74.27	75.67	73.72	**75.71**
bottle	79.38	79.18	79.67	**86.23**
bus	91.49	91.37	**93.99**	93.78
car	88.31	88.91	89.59	**90.02**
cat	87.11	89.61	94.54	**94.94**
chair	35.15	36.65	**45.86**	44.60
cow	86.04	89.54	89.78	**90.18**
table	60.36	**61.17**	50.42	53.70
dog	86.93	87.23	90.21	**91.55**
horse	86.73	87.73	89.04	**89.56**
motorbike	85.96	85.26	88.36	**88.80**
person	85.27	86.27	**89.04**	89.00
plant	61.23	61.43	59.04	**63.89**
sheep	86.21	87.21	87.34	**89.51**
sofa	47.19	47.70	**52.38**	49.69
train	84.18	84.38	87.47	**87.99**
television	75.47	76.37	75.36	**79.29**
mIoU(%)	77.59	78.55	78.89	**80.14**

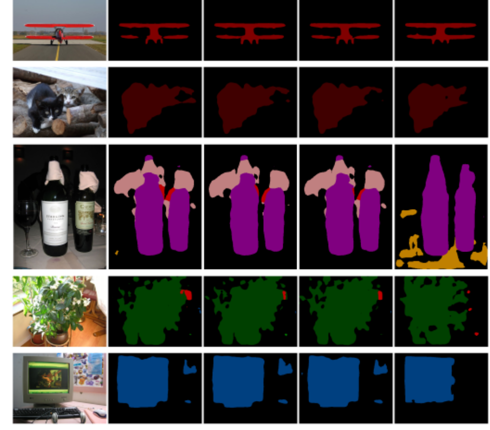


图 6 Pascal VOC 2012 测试集语义分割结果

Fig. 6 Semantic segmentation results of Pascal VOC 2012 testing set

3 结束语

本文提出一种基于卷积神经网络的语义分割方法，通过本文提出的联合特征金字塔模型(JFP)融合残差网络的三层输出，更加完整地提取图像特征，结合 ASPP 模型进一步提取图像特征，设计了一个简单的解码结构恢复图像尺寸，在此之外又设计了一个注意力模型作为辅助网络，辅助语义分割网络进行训练，本文的方法解决了特征提取信息丢失和网络训练收敛慢的问题。最后在 Pascal VOC 2012 数据集上的对比结果表示，本文提出的方法在 mIoU 上相比 Context+Decoder+CRFs[22]、MsNet-4[21]和 Auto-Deeplab[18]三种方法提高了将近 5%，相比 LadderDenseNet[17]、DeeplabV3[11]、DeeplabV3+[12]和 SDN[16]四种方法，本文使用增强数据集时，mIoU 有 1%～2%的提升，同时也超过 DFN[20]、PAN[13]和 DUpsamling[19]这三种方法近 0.5%。未来对于图像语义分割的工作是寻求更优的方法，设计一个更加优化的模型，提取图像的细节特征，进一步提高语义分割的效果。

参考文献：

[1] Gao Hongyang, Yuan Hao, Wang Zhengyang, *et al.* Pixel Deconvolutional Networks [EB/OL]. [2017-11-27]. https://arxiv.org/abs/1705. 06820.

[2] 蒋应锋, 张桦, 薛彦兵, 等. 一种新的多尺度深度学习图像语义理解方法研究 [J]. 光电子·激光, 2016, 27 (02): 224-230. (Jiang Yingfeng, Zhang Hua, Xue Yanbing, *et al.* Research on a new method of multi-scale deep learning image semantic understanding [J]. Optoelectronics laser, 2016, 27 (02): 224-230.)

[3] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [C]// Proc of the Conference on Computer Vision and Pattern Recognition: IEEE Press, 2015: 3431-3440.

[4] 董晓亚, 赵晓丽, 张嘉祺. 一种改进的噪声图像语义分割方法 [J]. 光电子·激光, 2017, 28 (12): 1372-1377. (Dong Xiaoya, Zhao Xiaoli, Zhang Jiaqi. An improved semantic segmentation method of noise image [J]. Optoelectronics laser, 2017, 28 (12): 1372-1377.)

[5] Li Jingwei, Yang Hua, Chen Lin, *et al.* Image semantic segmentation optimization by Conditional Random Field integrated with object clique potential [C]// Proc of the International Symposium on Broadband Multimedia Systems and Broadcasting: IEEE Press, 2017. 1-6.

[6] Badrinarayanan V, Kendall A, Cipolla R, *et al.* SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.

[7] Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation [C]// Proc of the Visual Communications and Image Processing: IEEE Press, 2017. 1-4.

[8] 李琳辉, 钱波, 连静, 等. 基于卷积神经网络的交通场景语义分割方法研究 [J]. 通信学报, 2018, 39 (04): 123-130. (Li Linhui, Qian Bo, Lian Jing, *et al.* Research on traffic scene semantic segmentation based on convolutional neural network [J]. Journal of communications, 2018, 39 (04): 123-130.)

[9] Lin Guosheng, Shen Chunhua, Hengel A V D, *et al.* Exploring Context with Deep Structured Models for Semantic Segmentation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (6): 1352-1366.

[10] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions [J]. arXiv: 151107122, 2015.

[11] Chen L-C, Papandreou G, Murphy K, *et al.* DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 834-848.

[12] Chen L-C, Zhu Y, Papandreou G, *et al.* Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [C]// Proc of the European Conference on Computer Vision: ECCV Press, 2018. 801-818.

[13] Li Hanchao, Xiong Pengfei, An Jie, *et al.* Pyramid Attention Network for Semantic Segmentation [EB/OL]. [2018-5-25]. https://arxiv.org/abs/1805. 10180.

[14] Yu F, Wang Dequan, Shelhamer E, *et al.* Deep Layer Aggregation [C]// Proc of the Conference on Computer Vision and Pattern Recognition: IEEE Press, 2018. 2403-2412.

[15] Dvornik N, Shmelkov K, Mairal J, *et al.* BlitzNet: A Real-Time Deep Network for Scene Understanding [C]// Proc of the International Conference on Computer Vision: IEEE Press, 2017. 4174-4182.

[16] Fu Jun, Liu Jing, Wang Yuhang, *et al.* Stacked Deconvolutional Network for Semantic Segmentation [EB/OL]. [2017-8-16]. https://arxiv.org/abs/1708. 04943.

[17] Krapac J, Segvic I K S. Ladder-Style DenseNets for Semantic

Segmentation of Large Natural Images [C]// Proc of the International Conference on Computer Vision Workshops: IEEE Press, 2017. 238-245.

[18] Liu Chenxi, Chen L-C, Schroff F, *et al.* Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation [C]// Proc of the Conference on Computer Vision and Pattern Recognition: IEEE Press, 2019. 82-92.

[19] Tian Zhi, He Tong, Shen Chunhua, *et al.* Decoders Matter for Semantic Segmentation: Data-Dependent Decoding Enables Flexible Feature Aggregation [EB/OL]. [2019-03-05]. https://arxiv. org/abs/1903. 02120.

[20] Yu Changqian, Wang Jingbo, Peng Chao, *et al.* Learning a Discriminative Feature Network for Semantic Segmentation [C]// Proc of the Conference on Computer Vision and Pattern Recognition: IEEE Press, 2018. 1857-1866.

[21] 陈智. 基于卷积神经网络的语义分割研究 [D]. 北京交通大学, 2018. (Chen Zhi. Research on semantic segmentation based on convolutional neural network [D]. Beijing Jiaotong University, 2018.)

[22] 孙海川. 基于全卷积网络的图像语义分割算法研究 [D]. 哈尔滨工业大学, 2018. (Sun Haichuan. Research on image semantic segmentation algorithm based on full convolution network [D]. Harbin University of technology, 2018.)